nature genetics

ARTICLES

https://doi.org/10.1038/s41588-020-0599-0

Check for updates

Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia

Pirro G. Hysi[1,2,3,26,27] ✉, Hélène Choquet[4,26], Anthony P. Khawaja[5,6,26], Robert Wojciechowski[7,8,26], Milly S. Tedja[9,10,26], Jie Yin[4], Mark J. Simcoe[2], Karina Patasova[1], Omar A. Mahroo[1,5], Khanh K. Thai[4], Phillippa M. Cumberland[3,11], Ronald B. Melles[12], Virginie J. M. Verhoeven[9,10,13], Veronique Vitart[14], Ayellet Segre[15], Richard A. Stone[16], Nick Wareham[6], Alex W. Hewitt[17], David A. Mackey[17,18], Caroline C. W. Klaver[9,10,19,20], Stuart MacGregor[21], The Consortium for Refractive Error and Myopia[25], Peng T. Khaw[5], Paul J. Foster[5,22], The UK Eye and Vision Consortium[25], Jeremy A. Guggenheim[23], 23andMe Inc.[25], Jugnoo S. Rahi[3,5,11,24,27], Eric Jorgenson[4,27] and Christopher J. Hammond[1,2,27]

Refractive errors, in particular myopia, are a leading cause of morbidity and disability worldwide. Genetic investigation can improve understanding of the molecular mechanisms that underlie abnormal eye development and impaired vision. We conducted a meta-analysis of genome-wide association studies (GWAS) that involved 542,934 European participants and identified 336 novel genetic loci associated with refractive error. Collectively, all associated genetic variants explain 18.4% of heritability and improve the accuracy of myopia prediction (area under the curve (AUC) = 0.75). Our results suggest that refractive error is genetically heterogeneous, driven by genes that participate in the development of every anatomical component of the eye. In addition, our analyses suggest that genetic factors controlling circadian rhythm and pigmentation are also involved in the development of myopia and refractive error. These results may enable the prediction of refractive error and the development of personalized myopia prevention strategies in the future.

Refractive errors occur when converging light rays from an image do not clearly focus on the retina. They are the seventh most prevalent clinical condition[1] and the second leading cause of disability in the world[2]. The prevalence of refractive error is rapidly increasing, mostly driven by a dramatic rise in the prevalence of one of its forms, myopia (near-sightedness). Although the causes of such a rise over a short time are probably due to environmental and cultural changes from the mid-twentieth century[3], refractive errors are highly heritable[4]. Several studies[5,6] have previously sought to identify genes controlling molecular mechanisms that lead to refractive error and myopia. However, the variance and heritability that can be attributed to known genetic factors is modest[7] and our knowledge of pathogenic mechanisms remains partial. Here, we conduct a meta-analysis that combines data from quantitative spherical equivalent and myopia status from large and previously unpublished GWAS of more than half a million subjects from

[1]Section of Ophthalmology, School of Life Course Sciences, King's College London, London, UK. [2]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [3]UCL Great Ormond Street Institute of Child Health, University College London, London, UK. [4]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [5]NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. [6]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK. [7]Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA. [8]Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD, USA. [9]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands. [10]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [11]Ulverscroft Vision Research Group, UCL Great Ormond Street Institute of Child Health, University College London, London, UK. [12]Department of Ophthalmology Kaiser Permanente Northern California, Redwood City, CA, USA. [13]Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands. [14]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, UK. [15]Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear, Boston, MA, USA. [16]Department of Ophthalmology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [17]Department of Ophthalmology, Royal Hobart Hospital, Hobart, Tasmania, Australia. [18]Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, Western Australia, Australia. [19]Department of Ophthalmology, Radboud University Medical Center, Rotterdam, the Netherlands. [20]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland. [21]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [22]Division of Genetics and Epidemiology, UCL Institute of Ophthalmology, London, UK. [23]School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK. [24]Department of Ophthalmology and NIHR, Biomedical Research Centre, Great Ormond Street Hospital NHS Foundation Trust, London, UK. [25]A list of members and affiliations appears in the Supplementary Note. [26]These authors contributed equally: Pirro G. Hysi, Hélène Choquet, Anthony P. Khawaja, Robert Wojciechowski, Milly S. Tedja. [27]These authors jointly supervised this work: Pirro G. Hysi, Jugnoo S. Rahi, Eric Jorgenson, Christopher J. Hammond. ✉e-mail: pirro.hysi@kcl.ac.uk

the UK Biobank, 23andMe and the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohorts, with subsequent replication and meta-analysis with data that were previously reported from the Consortium for Refractive Error and Myopia (CREAM).

Results

Association results. Analyses were restricted to subjects of European ancestry (Extended Data Fig. 1) and combined results from quantitative measures of spherical equivalent and categorical myopia status. Spherical equivalent quantifies refractive error; a negative spherical equivalent below a certain threshold defines myopia. We used results obtained from GWAS of directly measured spherical equivalent in 102,117 population-based UK Biobank participants[8] and 34,998 subjects participating in the GERA Study[9], and combined them with the results of analyses of self-reported myopia in 106,086 cases and 85,757 controls from the customer base of 23andMe, a personal genomics company[10]. Additionally, we included results from an analysis on the refractive status inferred using demographic and self-reported information on age at first use of prescription glasses among the UK Biobank participants not contributing to the quantitative GWAS (108,956 likely myopes to 70,941 likely non-myopes, see Methods). All analyses were adjusted for age, sex and main principal components. To obtain an overall association with refractive error, we meta-analyzed the results from all studies by using the z scores from the GWAS of the spherical equivalent and the negative values of z scores from the case–control studies (23andMe and UK Biobank), as myopia is negatively correlated with spherical equivalent. As expected, the large total sample size of the discovery meta-analysis ($n=508{,}855$) led to a nominally large genomic inflation factor ($\lambda=1.94$). The linkage disequilibrium (LD) score regression intercept was (1.17), and the (intercept-1)/(mean (X^2) − 1) ratio of 0.097 is fully in line with the expectations of polygenicity[11].

We found associations for 438 discrete genomic regions (Fig. 1 and Supplementary Table 1), defined by markers that were contiguously associated at the conventional level of GWAS significance[12,13] of $P<5\times10^{-8}$, separated by more than 1 Mb from other GWAS-associated markers, as recommended elsewhere[14]. Among these, 308 loci, including 14 on the X chromosome, were not described in previous GWAS studies of refractive error[7]. The observed effect sizes were consistent across all of the studies (Supplementary Table 1 and Supplementary Data Set 1). The association with refractive error was statistically strongest for rs12193446 ($P=9.87\times10^{-328}$), within *LAMA2*, a gene that has been associated previously with refractive error[5,6] and in which mutations cause muscular dystrophy[15]. Consistent with these *LAMA2* properties, polymorphisms located within the genes that encode both major *LAMA2* receptors, *DAG1* (ref. [16]) ($P=1.67\times10^{-8}$ for rs111327216) and *ITGA7* (ref. [17]) ($P=8.57\times10^{-9}$ for rs17117860), which are also known causes of muscular dystrophy[18,19], were significantly associated with refractive error in the discovery meta-analysis.

We compared our discovery meta-analysis findings with GWAS results from 34,079 participants in the CREAM consortium, who were part of a previously reported meta-analysis[7]. To avoid any potential overlap with the UK Biobank participants, only samples from non-UK European CREAM participants were used for replication. Despite the vast power differential, 55 of the SNPs that showed the strongest association in their respective regions in the discovery meta-analysis were significant after Bonferroni correction in the replication sample. A further 142 had a false discovery rate (FDR) < 0.05 and 192 were nominally significant at $P<0.05$ (Supplementary Table 2). The effect sizes observed in the discovery and replication samples were strongly correlated (Pearson's $r=0.91$; Extended Data Fig. 2). Meta-analysis of all five cohorts (discovery and replication) expanded the number to 449 associated regions of variable length and number of SNPs (Extended Data Fig. 3), of which 336 regions were novel (Supplementary Table 3).

Most of the 449 refractive-error-associated regions contained at least one gene that was linked to severe ocular manifestations in the Online Mendelian Inheritance In Man (OMIM, https://omim.org) resource or other genes with interesting links to eye disease (Supplementary Table 4). Although most loci that were identified through our meta-analyses were novel, several of them hosted genes that harbor mutations leading to myopia or other refractive error phenotypes (Supplementary Data 2). Several genes that were found to be significantly associated with refractive error were linked to Mendelian disorders that affect corneal structure, some of which code for transcription factors involved in corneal development[20] (Supplementary Table 5). Mutations in these genes cause corneal dystrophies (*SLC4A11*, $P=5.81\times10^{-11}$ for rs41281858; *TCF4*, $P=4.14\times10^{-8}$ for rs41396445; *LCAT*, $P=1.26\times10^{-10}$ for rs5923 and *DCN*, $P=3.67\times10^{-9}$ for rs1280632), megalocornea (*LTBP2*, $P=1.91\times10^{-24}$ for rs73296215) and keratoconus (*FNDC3B*, $P=1.89\times10^{-14}$ for rs199771582, which was described previously[7]). Eleven refractive-error-associated genes were linked to anomalies of the crystalline lens (Supplementary Table 6), including genes linked to autosomal dominant cataracts (*PAX6*, which was previously linked to myopia[21], $P=8.31\times10^{-11}$ for rs1540320; *PITX3*, $P=1.05\times10^{-10}$ for rs7923183; *MAF*, $P=5.50\times10^{-9}$ for rs16951312; *CHMP4B*, $P=9.95\times10^{-11}$ for rs6087538; *TDRD7*, $P=4.79\times10^{-8}$ for rs13301794) and lens ectopia (*FBN1*, $P=3.30\times10^{-24}$ for rs2017765; *ADAMTSL4*, $P=8.19\times10^{-14}$ for rs12131376). Some of the genes affected several eye components. For example, *LTBP2* variants are also associated with congenital glaucoma[22], and *COL4A3* (rs7569375, $P=1.14\times10^{-8}$) causes Alport syndrome, which manifests with abnormal lens shape (lenticonus) and structural changes in the retina.

Association was also observed within or near 13 genes that are known to harbor mutations that cause microphthalmia (Supplementary Table 7), including *TENM3* ($P=2.48\times10^{-11}$ for rs35446926); *OTX2* ($P=6.15\times10^{-11}$ for rs928109); *VSX2* ($P=4.60\times10^{-10}$ for rs35797567); *MFRP* ($P=2.85\times10^{-16}$ for rs10892353); and the previously identified[6] *TMEM98* ($P=3.49\times10^{-43}$, for rs62067167). Association was also identified for *VSX1* ($P=4.59\times10^{-8}$ for rs6050351), a gene that is closely regulated by VSX2 (ref. [23]) and believed to have important roles in eye development[24]. Many of the genes nearest to the associated SNPs have been linked to inherited retinal disease (Supplementary Table 8), including 32 genes linked to cone–rod dystrophies, night blindness and retinitis pigmentosa and age-related macular degeneration (*HTRA1/ARMS2*). Among genes that were identified in novel regions and are associated with refractive error, *ABCA4* ($P=3.20\times10^{-10}$ for rs11165052) and *HTRA1/ARMS2* ($P=5.72\times10^{-23}$ for rs2142308) are linked to macular disorders, and numerous others, such as *FBN2* ($P=8.63\times10^{-11}$ for rs6860901), *TRAF3IP1* ($P=5.71\times10^{-16}$ for rs7596847) and CWC27 ($P=1.84\times10^{-18}$ for rs1309551) are linked to retinitis pigmentosa, retinal dystrophy and other retinal diseases. Significant association was identified near other genes of interest such as *DRD1* ($P=4.51\times10^{-16}$ for rs13190379), a dopamine receptor. Together, these results are consistent with previous suggestions of light transmission and transduction as potential mechanisms in refractive error[7,25].

Wnt signaling has previously been implicated in experimental myopia[26]. We found significant association near several genes that encode Wnt proteins (*WNT7B*, a gene previously associated with axial length[27], $P=1.42\times10^{-26}$ for rs73175083; *WNT10A*, previously associated with central corneal thickness[28], $P=1.65\times10^{-17}$ for rs121908120 and *WNT3B*, $P=8.52\times10^{-16}$ for rs70600), which suggests that organogenesis through Wnt signaling is likely to be involved in refractive error. Significant association was identified at genes that encode key members of the canonical Wnt pathway (for example, *CTNNB1*, $P=7.30\times10^{-27}$ for rs13072632 and *AXIN2*, $P=1.40\times10^{-8}$ for rs9895291) and of the non-canonical Wnt

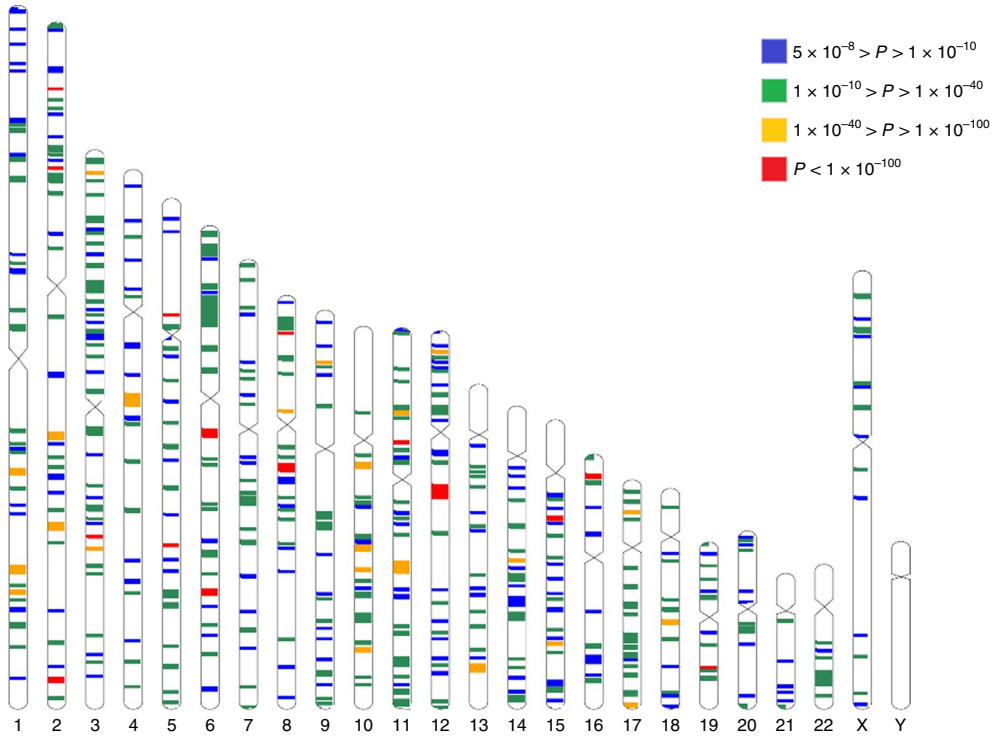


Fig. 1 | All GWAS-associated regions from the main meta-analysis. Each band is a true scale of the genomic regions associated with refractive error that are listed in Supplementary Table 1 (+250 kb on each side to make smaller regions more visible). The different color codes represent the significance (*P* value) for the genetic variant within that region that displays the strongest evidence for association.

pathway (for example, *NFATC3*, $P = 1.493 \times 10^{-12}$ for rs147561310), or at genes that are known to encode members of both pathways (for example, *RHOA*, $P = 1.81 \times 10^{-11}$ for rs7623687, or the previously described[7] *TCF7L2* gene, $P = 9.38 \times 10^{-46}$ for rs56299331; Supplementary Table 9).

Similarly to results reported from previous published analyses[25], we found associations for genes that are involved in the transport of sodium, potassium, calcium, magnesium and other cations (Supplementary Table 10). The involvement of genes related to glutamatergic synaptic transmission was also notable (Supplementary Table 11). Glutamate is one of the main transmitters in the first synapse to be released by photoreceptors toward bipolar cells and is the main excitatory neurotransmitter of the retina, and expression of genes that participate in glutamate signaling pathways is significantly altered in myopia models[29]. These associations support the involvement in refractive error pathogenesis of neurotransmission and neuronal depolarization and hyperpolarization, which were also suggested before[7]. Associations with *POU6F2* gene intronic variants (rs2696187, $P = 1.11 \times 10^{-11}$) also suggest involvement of factors related to development of amacrine and ganglion cells[30]. Other genes at refractive-error-associated loci were annotated in OMIM to infantile epilepsy, microcephaly, severe learning difficulty or other inborn diseases that affect the central nervous system (CNS) (Supplementary Table 12).

Polymorphisms in genes that are linked to oculocutaneous albinism (OCA) were significantly associated with refractive error (Supplementary Table 13), although typically association was identified for SNPs that were not strongly associated with other pigmentation traits[31]. Strong association with refractive error was identified near the *OCA2* gene, which causes OCA type 2 ($P = 1.37 \times 10^{-15}$ for rs79406658), and in genes that are linked to OCA type 3 (*TYRP1*, $P = 1.18 \times 10^{-11}$ for rs62538956), OCA type 5 (*SLC39A8*, $P = 4.03 \times 10^{-17}$ for rs13107325) and OCA type 7 (*LRMDA* (formerly known as *C10orf11*), $P = 1.73 \times 10^{-16}$ for rs12256171). In addition, significant association was identified near genes that are linked to ocular albinism on chromosome X (*TBL1X* and *GPR143* (ref. [32]), $P = 2.20 \times 10^{-18}$ for rs34437079) and Hermansky–Pudlak Syndrome albinism (*BLOC1S1*, $P = 2.46 \times 10^{-22}$, for rs80340147; note that this gene forms a conjoined read-through transcript with *RDH5*; *BLOC1S1–RDH5*). Other associated markers were located within genes that are involved in systemic pigmentation, which has also been associated previously with refractive error[7], such as *RALY* ($P = 3.14 \times 10^{-18}$ for rs2284388) and *TSPAN10* ($p = 2.22 \times^{-50}$, rs9747347), as well as in melanoma (*MCHR2*, $P = 2.37 \times 10^{-15}$ for rs4839756).

Functional properties of the associated markers. Among the significantly associated markers, 367 unique markers were frameshift or missense variants (Supplementary Table 14). Several are nonsynonymous, such as rs1048661 within *LOXL1* (a gene that causes pseudoexfoliation syndrome and glaucoma[33]) that results in a p.Arg141Leu alteration, and rs10490924 in *ARMS2* (associated with increased susceptibility to age-related macular degeneration[34]) that results in a p.Ala69Ser alteration. Other associated variants with predicted deleterious consequences were located in several genes, such as *RGR* ($P = 6.89 \times 10^{-68}$ for rs1042454), a gene previously associated with refractive error[7,10] and retinitis pigmentosa[35], and within the *FBN1* gene, near clusters of mutations that cause Marfan syndrome and anterior segment dysgenesis[36].

Because the functional link between other associated variants and development of refractive error phenotypes is less obvious, we next performed gene-set enrichment analyses to identify properties that are significantly shared by genes that were identified by the final meta-analysis of all cohorts. An enrichment analysis of Gene Ontology processes (Supplementary Table 15) identified enrichment for genes that participate in RNA polymerase II transcription regulation ($P = 1 \times 10^{-6}$) and nucleic acid binding transcription factor activity ($P = 1.10 \times 10^{-6}$), which suggests that many

of the genetic associations we identified interfere with gene expression. The genes that were identified by the meta-analyses were also significantly enriched in the Gene Ontology terms 'eye development' ($P=6.10\times10^{-6}$) and 'circadian regulation of gene expression' ($P=1.10\times10^{-4}$).

A transcription factor binding site enrichment analysis identified significant (FDR < 0.05) over-representation of sites targeted by *GATA4*, *EP300* and *RREB1*, for which association was observed in the meta-analyses (Supplementary Table 16). Binding sites of transcription factors involved in eye morphogenesis and development such as *MAF* (in which mutations cause autosomal cataract), *FOXC1* and *PITX2* (in which mutations cause anterior segment dysgenesis) or *CRX* (in which mutations cause cone–rod dystrophy) were also enriched. Binding sites for *CRX* and *PAX4* were also significantly enriched; these transcription factors are two of the regulators of circadian rhythm and melatonin synthesis[37] alongside *OTX2*, for which significant SNP association was observed in our refractive error meta-analysis ($P=6.15\times10^{-11}$ for rs928109). All of these enriched gene sets were observed for the first time in a GWAS analysis, although some of the mechanisms that relate them to refractive error and myopia have been proposed previously[38].

Many of the variants that were associated with refractive error in our analyses are located within or near genes that are expressed in numerous body tissues (Extended Data Fig. 4), and in particular in the nervous system, which is consistent with our evidence of extraocular, CNS involvement in refractive error. Within the eye, these genes were particularly strongly expressed in eye tissues such as cornea, ciliary body, trabecular meshwork[39] and retina[40] (Extended Data Fig. 5 and Supplementary Table 17). A stratified LD score regression applied to specifically expressed genes[41] revealed that the results of the GWAS were most strongly correlated with genes that are expressed in the retina and basal ganglia in the CNS but these correlations were not significant after multiple testing correction (Extended Data Fig. 6 and Supplementary Table 18). It is possible that the strength of these correlations was constrained by the fact that in most cases, available expression levels were measured in adult samples, whereas refractive error and myopia primarily develop at younger ages.

A summary data-based Mendelian randomization (SMR) analysis[42], which integrates GWAS data with eQTL data from peripheral blood[43] and brain tissues[44], identified concomitant association with refractive error and eQTL transcriptional regulation effects for 159 and 97 genes, respectively (Supplementary Tables 19,20). A similar analysis that integrates GWAS summary data with methylation data from brain tissues identified association with both refractive error and changes in methylation for 134 genes (Supplementary Table 21).

Genetic effects shared between refractive error and other conditions. Examination of the GWAS Catalog[45], revealed that some of the genetic variants reported here were previously associated with refractive error and with other traits, in particular intraocular pressure, intelligence and education; the latter two are known myopia risk factors (Supplementary Table 22). We used LD score regression to assess the correlation of genetic effects between refractive error and other phenotypes from GWAS summary statistics (Supplementary Table 23). Refractive error genetic risk was found to correlate significantly with intelligence, both in childhood[46] ($r_g=-0.27$, $P=4.76\times10^{-9}$) and in adulthood (fluid intelligence score $r_g=-0.25$, $P=1.56\times10^{-39}$), educational attainment (defined as the number of years spent in formal education; $r_g=-0.24$, $P=3.36\times10^{-54}$), self-reported cataract ($r_g=-0.31$, $P=4.70\times10^{-10}$) and intraocular pressure (IOP; $r_g=-0.14$, $P=1.04\times10^{-12}$).

Higher educational attainment appears to cause myopia as demonstrated by Mendelian randomization studies[47]. A gene-by-environment interaction GWAS for spherical equivalent and educational attainment (using age at completion of formal full-time education as a proxy) was conducted in 66,242 UK Biobank participants. Despite the relatively well-powered sample, only one locus yielded evidence of statistically significant interaction (rs536015141 within *TRPM1*, $P=2.35\times10^{-9}$; Supplementary Table 24), which suggests that the true relationship between refractive error and education is compounded by several factors and may not be linear in nature as was suggested recently[48]. The TRPM1 protein is localized in rod ON bipolar cell dendrites, and rare mutations in this gene cause congenital stationary night blindness[49], which is often associated with high myopia.

To further explore the nature of the relationship between refractive error and IOP, we built Mendelian randomization models using genetic effects that were reported previously for IOP[50]. On average, every 1 mmHg increase in IOP predicts a 0.05–0.09 diopter (D) decrease in spherical equivalent (Supplementary Table 25 and Extended Data Fig. 7). We also built a Mendelian randomization model to assess the relationship between intelligence and spherical equivalent, but statistical evidence in this case points toward genetic pleiotropy rather than causation (Supplementary Table 26). This suggests that both myopia and intelligence are often influenced by the same factors, but without a direct causal path linking one to the other. We found no significant genetic correlations between refractive error and the glaucoma endophenotype vertical cup-to-disc ratio ($r_g=-0.01$, $P=0.45$), or hair pigmentation ($r_g=-0.03$, $P=0.35$). Therefore, refractive error and pigmentation may have different allelic profiles with limited sharing of genetic risk.

Conditional analysis and risk prediction. We subsequently carried out a conditional analysis[51] on the meta-analysis summary results and found a total of 904 independent SNPs that were significantly associated with refractive error. Of these markers, 890 were available in the EPIC-Norfolk study, an independent cohort that did not participate in the refractive error meta-analysis (Extended Data Fig. 8). These markers alone explained 12.1% of the overall spherical equivalent phenotypic variance in a regression model or 18.4% (s.e. = 0.04) of the spherical equivalent heritability. Newly associated markers that were identified in our meta-analysis, but not in the previous large GWAS[7], explain 4.6% (s.e. = 0.01) of the spherical equivalent phenotypic variance in the EPIC-Norfolk study, which is an improvement of one-third compared to heritability explained by previously associated markers[7].

Predictive models that were based on the above-mentioned 890 SNPs, along with age and sex, were predictive of myopia (versus all non-myopia controls) with AUC values of 0.67, 0.74 and 0.75, respectively (Fig. 2), depending on the severity cutoff for myopia (≤−0.75 D, ≤−3.00 D and ≤−5.00 D, respectively). The performance of the predictions appears not to improve for myopia definitions of −3.00 D or worse, which suggests that the information extracted from our meta-analysis is more representative of the genetic risk for common myopia that is seen in the general population, than for more severe forms of myopia, which may have a distinctive genetic architecture.

Further exploration of refractive error genetic architecture. Using information from over half a million population-based participants SNPs identified in these analyses still explains only 18.4% of the spherical equivalent heritability. We next assessed how many common SNPs are likely to explain the entire heritable component of refractive error, and what sample sizes are likely to be needed in the future to identify them, using the likelihood-based approach that is described elswhere[52]. We estimate that approximately 13,808 (s.e. = 969) polymorphic variants are likely to be behind the full refractive error heritability. In a similar manner to the results of other quantitative phenotypic traits that have been published previously[52], our analyses estimate that 10.3% (s.e. = 1.0%) of the phenotypic variance is probably explained by a batch of approximately

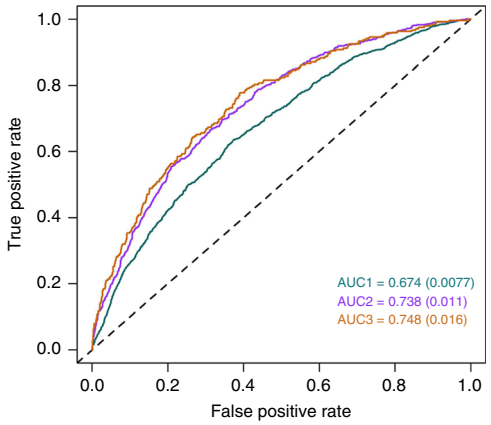


Fig. 2 | Receiver operating characteristic curves for myopia predictions, using information from 890 SNP markers identified in the meta-analysis. The three different colors represent three different curves for each of the different definitions of myopia: green, all myopia (defined as <−0.75 D); magenta, moderate myopia (<−3.00 D); and brown, severe myopia (<−5.00 D).

543 (s.e.=81) common genetic variants of relatively large effect size and a further 20.8% (s.e.=0.9%) of the entire phenotypic variance may be explained by the remainder. With increased sample sizes, we suggest that the proportion of variance that can be explained will continue to improve quickly but will begin to plateau for sample sizes above 1 million, after which further increases in sample size will probably yield ever-diminishing additional phenotypic variance (Extended Data Fig. 9).

Discussion

Our results provide evidence for at least two major sets of mechanisms in the pathogenesis of refractive error. The first affects intraocular pressure, eye structure, ocular development and physiology, and the second is CNS related and includes circadian rhythm control. Contributors to refractive error include all anatomical factors that alter refractive power relative to eye size, light transmittance, photoconductance and higher cerebral functions.

The findings implicate almost all anatomical components of the eye, which, along with the CNS, participate in the development of refractive error. The healthy cornea contributes to 70% of the optical refractive power of the eyes[53], and genes involved in corneal structure, topography and function may directly contribute to refractive error through direct changes in the corneal refraction. Our results show that several genes involved in lens development also contribute to refractive error in the general population. It is unclear whether their contribution is mediated through alterations in biomechanical properties that affect the ability of the eyes to accommodate, changes to the lens refractive index, or alterations in light transmission properties that impair the ability to focus images on the retina.

Many retinal genes are implicated in the development of refractive error, thus reflecting the role of light in mediating eye growth and the importance of the role of the retina in light transduction and processing[7]. Associations with refractive error at genes that encode gated ion channels and glutamate receptors point to the photoreceptor–bipolar cell interface as a potential key factor in refractive error. In several of the associated genes, rare mutations cause night blindness, thus primarily implicating the rod system in the pathophysiology of refractive error, although the same mutations are present and may also affect the cone-related pathways. The *TRPM1* gene, which is important for rod ON bipolar cell polarity[54], is also implicated in the gene–education interaction analysis. Associations observed for the *VSX1* gene and for VSX2, its negative regulator, implicate the cone bipolar cells[55].

The association with refractive error at genes involved in pigmentation, including most of the OCA-causing genes, raises questions about the relationship between melanin, pigmentation and eye growth and development. These associations are unlikely to be influenced by any cryptic population structure in our samples, which our analyses were designed to control. None of the major pigmentation-associated SNPs[31] were directly associated with refractive error and there was no significant correlation of genetic effects between refractive error and pigmentation.

The mechanisms that link pigmentation with refractive error are unclear. Foveal hypoplasia[56] and optic disc[57] dysplasias are common in all forms of albinism[58]. Although melanin synthesis is disrupted in albinism, both melanin and dopamine are synthesized through shared metabolic pathways. Disc and chiasmal lesions in albinism are often attributed to dopamine[59], but we found limited evidence to support an association with refractive error for genetic variants involved in dopamine signaling. The scarcity of association with refractive error for genes involved in dopamine-only pathways contrasts with the abundance of association for genes involved in pigmentation and melanin synthesis. This may suggest that melanin metabolism is connected to refractive error through other mechanisms that are independent from the metabolic pathways it shares with dopamine production. Melanin reaches the highest concentrations in the retinal pigment epithelium at the outmost layer of the retina and anteriorly in the iris, and variations in pigmentation may affect the intensity of the light that reaches the retina. Light exposure is a major protective factor for development of myopia[60,61]. It is possible that pigmentation has a role in light signal transmission and transduction.

Animal model experiments suggest that in addition to local ocular mechanisms, emmetropization (the process by which the eye develops to minimize refractive error) is strongly influenced by the CNS[62]. The strong correlation of genetic risks between refractive error and intelligence, and the association of genes that are linked to severe learning disability, support the involvement of the CNS in emmetropization and refractive error pathogenesis.

Results from gene-set enrichment analysis demonstrate an interesting evolution with increasing sample sizes. Although smaller previous studies were sufficiently powered to discover enrichment of low, cell-level properties, such as cation channel activity and participation in the synaptic space structures[25], recent studies with considerably more power have provided additional evidence for enrichment and involvement of more integrated physiological functions, such as light signal processing in retinal and other cells[7]. In addition to the identification of a much larger number of gene associations and the explanation of higher proportions of heritability, our results, which are derived from a sample with considerably more statistical power, uphold the previous findings and support the involvement of the same molecular and physiological mechanisms that were described previously.

In line with expectations of the greater power used in our study to discover associations with genes and gene sets that are individually responsible for smaller proportions of the refractive error variance[63], we found evidence for higher regulatory mechanisms that act more holistically on eye development or integrate eye growth and homeostasis with other processes of an extraocular nature. For example, we found evidence that binding sites of transcription factors involved in the control of circadian rhythm are significantly enriched among genes that are associated with refractive error. Circadian rhythm is important in emmetropization and its disruption leads to myopia in animal knockout models[38], potentially through dopamine-mediated mechanisms, or changes in IOP and diurnal variations.

Most of the loci that were identified through our meta-analysis are not subject to particularly strong and systematic evolutionary

pressures (Extended Data Fig. 10). The variability in minor allele frequencies that was observed across loci associated with refractive error may therefore be the result of genetic drift. However, given the variety of the different visual components whose disruptions can result in refractive error, this variability may also be the result of overall balancing forces that encourage high allelic diversity of genes involved in refractive error, thereby providing additional buffering capacity to absorb environmental pressures[48] or genetic disruptions on any of the individual components of the visual system.

Our results cast light on potential mechanisms that contribute to refractive error in the general population and have identified the genetic factors that explain a considerable proportion of the heritability and phenotypic variability of refractive error. This allows us to substantially improve our ability to make predictions of myopia risk and generate novel hypotheses on how multiple aspects of visual processing affect emmetropization, which may pave the way to personalized risk management and treatment of refractive error in the population in the future.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0599-0.

Received: 24 March 2019; Accepted: 24 February 2020;
Published online: 30 March 2020

References

1. Vos, T. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1211–1259 (2017).
2. *The Global Burden of Disease: 2004 Update* (World Health Organization, 2008).
3. Williams, K. M. et al. Increasing prevalence of myopia in Europe and the impact of education. *Ophthalmology* **122**, 1489–1497 (2015).
4. Sanfilippo, P. G., Hewitt, A. W., Hammond, C. J. & Mackey, D. A. The heritability of ocular traits. *Surv. Ophthalmol.* **55**, 561–583 (2010).
5. Kiefer, A. K. et al. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genet.* **9**, e1003299 (2013).
6. Verhoeven, V. J. et al. Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat. Genet.* **45**, 314–318 (2013).
7. Tedja, M. S. et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat. Genet.* **50**, 834–848 (2018).
8. Cumberland, P. M. et al. Frequency and distribution of refractive error in adult life: methodology and findings of the UK Biobank study. *PLoS One* **10**, e0139780 (2015).
9. Kvale, M. N. et al. Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* **200**, 1051–1060 (2015).
10. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
11. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
12. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227–234 (2008).
13. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
14. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
15. Oliveira, J. et al. *LAMA2* gene analysis in a cohort of 26 congenital muscular dystrophy patients. *Clin. Genet.* **74**, 502–512 (2008).
16. Colognato, H. et al. Identification of dystroglycan as a second laminin receptor in oligodendrocytes, with a role in myelination. *Development* **134**, 1723–1736 (2007).
17. Burkin, D. J. & Kaufman, S. J. The α7β1 integrin in muscle development and disease. *Cell Tissue Res.* **296**, 183–190 (1999).
18. Ervasti, J. M. & Campbell, K. P. Dystrophin-associated glycoproteins: their possible roles in the pathogenesis of Duchenne muscular dystrophy. *Mol. Cell Biol. Hum. Dis. Ser.* **3**, 139–166 (1993).
19. Mayer, U. et al. Absence of integrin α7 causes a novel form of muscular dystrophy. *Nat. Genet.* **17**, 318–323 (1997).
20. Jean, D., Ewan, K. & Gruss, P. Molecular regulators involved in vertebrate eye development. *Mech. Dev.* **76**, 3–18 (1998).
21. Hammond, C. J., Andrew, T., Mak, Y. T. & Spector, T. D. A susceptibility locus for myopia in the normal population is linked to the *PAX6* gene region on chromosome 11: a genomewide scan of dizygotic twins. *Am. J. Hum. Genet.* **75**, 294–304 (2004).
22. Ali, M. et al. Null mutations in *LTBP2* cause primary congenital glaucoma. *Am. J. Hum. Genet.* **84**, 664–671 (2009).
23. Clark, A. M. et al. Negative regulation of Vsx1 by its paralog Chx10/Vsx2 is conserved in the vertebrate retina. *Brain Res.* **1192**, 99–113 (2008).
24. Heon, E. et al. *VSX1*: a gene for posterior polymorphous dystrophy and keratoconus. *Hum. Mol. Genet.* **11**, 1029–1036 (2002).
25. Hysi, P. G. et al. Common mechanisms underlying refractive error i dentified in functional analysis of gene lists from genome-wide association study results in 2 European British cohorts. *JAMA Ophthalmol.* **132**, 50–56 (2014).
26. Ma, M. et al. Wnt signaling in form deprivation myopia of the mice retina. *PLoS One* **9**, e91086 (2014).
27. Miyake, M. et al. Identification of myopia-associated *WNT7B* polymorphisms provides insights into the mechanism underlying the development of myopia. *Nat. Commun.* **6**, 6689 (2015).
28. Cuellar-Partida, G. et al. *WNT10A* exonic variant increases the risk of keratoconus by decreasing corneal thickness. *Hum. Mol. Genet.* **24**, 5060–5068 (2015).
29. Stone, R. A. et al. Image defocus and altered retinal gene expression in chick: clues to the pathogenesis of ametropia. *Invest. Ophthalmol. Vis. Sci.* **52**, 5765–5777 (2011).
30. Zhou, H., Yoshioka, T. & Nathans, J. Retina-derived POU-domain factor-1: a complex POU-domain gene implicated in the development of retinal ganglion and amacrine cells. *J. Neurosci.* **16**, 2261–2274 (1996).
31. Hysi, P. G. et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* **50**, 652–656 (2018).
32. Fabian-Jessing, B. K. et al. Ocular albinism with infertility and late-onset sensorineural hearing loss. *Am. J. Med. Genet. A* **176**, 1587–1593 (2018).
33. Thorleifsson, G. et al. Common sequence variants in the *LOXL1* gene confer susceptibility to exfoliation glaucoma. *Science* **317**, 1397–1400 (2007).
34. Rivera, A. et al. Hypothetical *LOC387715* is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum. Mol. Genet.* **14**, 3227–3236 (2005).
35. Morimura, H., Saindelle-Ribeaudeau, F., Berson, E. L. & Dryja, T. P. Mutations in *RGR*, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nat. Genet.* **23**, 393–394 (1999).
36. Robinson, P. N. et al. Mutations of *FBN1* and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. *Hum. Mutat.* **20**, 153–161 (2002).
37. Rohde, K., Moller, M. & Rath, M. F. Homeobox genes and melatonin synthesis: regulatory roles of the cone-rod homeobox transcription factor in the rodent pineal gland. *Biomed. Res. Int.* **2014**, 946075 (2014).
38. Chakraborty, R. et al. Circadian rhythms, refractive development, and myopia. *Ophthalmic Physiol. Opt.* **38**, 217–245 (2018).
39. Carnes, M. U., Allingham, R. R., Ashley-Koch, A. & Hauser, M. A. Transcriptome analysis of adult and fetal trabecular meshwork, cornea, and ciliary body tissues by RNA sequencing. *Exp. Eye Res.* **167**, 91–99 (2018).
40. Ratnapriya, R. et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).
41. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
42. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
43. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
44. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
45. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
46. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with *FNBP1L*. *Mol. Psychiatry* **19**, 253–258 (2014).

47. Mountjoy, E. et al. Education and myopia: assessing the direction of causality by mendelian randomisation. *BMJ* **361**, k2022 (2018).
48. Pozarickij, A. et al. Quantile regression analysis reveals widespread evidence for gene-environment or gene-gene interactions in myopia development. *Commun. Biol.* **2**, 167 (2019).
49. Audo, I. et al. *TRPM1* is mutated in patients with autosomal-recessive complete congenital stationary night blindness. *Am. J. Hum. Genet.* **85**, 720–729 (2009).
50. Khawaja, A. P. et al. Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. *Nat. Genet.* **50**, 778–782 (2018).
51. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
52. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
53. Zadnik, K. et al. Normal eye growth in emmetropic schoolchildren. *Optom. Vis. Sci.* **81**, 819–828 (2004).
54. Li, Z. et al. Recessive mutations of the gene *TRPM1* abrogate ON bipolar cell function and cause complete congenital stationary night blindness in humans. *Am. J. Hum. Genet.* **85**, 711–719 (2009).
55. Chow, R. L. et al. *Vsx1*, a rapidly evolving *paired*-like homeobox gene expressed in cone bipolar cells. *Mech. Dev.* **109**, 315–322 (2001).
56. Struck, M. C. Albinism: update on ocular features. *Curr. Ophthalmol. Rep.* **3**, 232–237 (2015).
57. Mohammad, S. et al. Characterization of abnormal optic nerve head morphology in albinism using optical coherence tomography. *Invest. Ophthalmol. Vis. Sci.* **56**, 4611–4618 (2015).
58. Yahalom, C. et al. Refractive profile in oculocutaneous albinism and its correlation with final visual outcome. *Br. J. Ophthalmol.* **96**, 537–539 (2012).
59. Lopez, V. M., Decatur, C. L., Stamer, W. D., Lynch, R. M. & McKay, B. S. L-DOPA is an endogenous ligand for OA1. *PLoS Biol.* **6**, e236 (2008).
60. Karouta, C. & Ashby, R. S. Correlation between light levels and the development of deprivation myopia. *Invest. Ophthalmol. Vis. Sci.* **56**, 299–309 (2014).
61. Wu, P. C., Tsai, C. L., Wu, H. L., Yang, Y. H. & Kuo, H. K. Outdoor activity during class recess reduces myopia onset and progression in school children. *Ophthalmology* **120**, 1080–1085 (2013).
62. Troilo, D., Gottlieb, M. D. & Wallman, J. Visual deprivation causes myopia in chicks with optic nerve section. *Curr. Eye Res.* **6**, 993–999 (1987).
63. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).

Methods

Study participants. *The UK Biobank.* The UK Biobank is a multisite cohort study of UK residents aged 40 to 69 years who were registered with the National Health Service (NHS) and living up to 25 miles from a study center. Detailed study protocols are available online (http://www.ukbiobank.ac.uk/resources/ and http://biobank.ctsu.ox.ac.uk/crystal/docs.cgi). The study was conducted with the approval of the North-West Research Ethics Committee (ref. 06/MRE08/65), in accordance with the principles of the Declaration of Helsinki, and all participants gave written informed consent.

Two separate groups of UK Biobank participants were included in these analyses. The first included participants whose refractive error was measured directly (non-cycloplegic autorefraction using the Tomey RC 5000 Auto-Refractor Keratometer). Direct measurements of refractive errors were available for 22.7% of the UK Biobank sample. To ensure reliable and accurate refractive error data, previously published quality control criteria were applied[8]. The spherical equivalent was calculated as spherical refractive error (UK Biobank codes 5084 and 5085) plus half the cylindrical error (UK Biobank 5086 and 5087) for each eye.

The second UK Biobank group included participants whose refractive error was not measured directly. The refractive error status of these participants was inferred using a questionnaire and other indirect data. Available demographic and clinical information were used to obtain an estimate about the likely myopia status of each individual. A support vector machine (SVM) model, with age, sex, age of first spectacle wear and year of birth as prediction parameters was used to infer the myopia status of participants. The phenotypic estimation was evaluated through a model that was initially trained in 80% of the randomly selected UK Biobank participants of European descent for whom direct spherical equivalent and refractive error status were available. Then the performance was assessed in the remaining 20% of UK Biobank participants of European descent for whom direct spherical equivalent and refractive error status were available. Finally, the SVM predictions in the remaining individuals with no direct spherical error measurements available using the model developed for the training data.

All UK Biobank genotypes were obtained as described elsewhere[64]. The UK Biobank team then performed imputation from a combined Haplotype Reference Consortium (HRC) and UK10K reference panel. Phasing on the autosomes was carried out using a modified version of the SHAPEIT2 (ref. [65]) program modified to allow for very large sample sizes. Only HRC-imputed variants were used for the purpose our analyses of the UK Biobank participants. The variant-level quality control exclusion metrics that were applied to imputed data for GWAS included the following: call rate < 95%, Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$, posterior call probability < 0.9, imputation quality < 0.4 and MAF < 0.005. The Y chromosome and mitochondrial genetic data were excluded from this analysis. In total, 10,263,360 imputed DNA sequence variants were included in our analysis. Participants of non-European ancestry and participants with relatedness corresponding to third-degree relatives or closer, as well as samples with an excess of missing genotype calls or heterozygosity were excluded. In total, genotypes were available for 102,117 participants of European ancestry with spherical equivalent data.

Association models in the first UK Biobank subset used as predictors the average of spherical equivalent as the outcome and allele dosages at each genetic locus. Mixed linear regressions were used, adjusted for age, sex and the first ten principal components, and were implemented in the Bolt-LMM software[66].

For the second UK Biobank subset, for which no direct spherical equivalent measurement was available, the mixed linear model was built with the predicted myopia status as outcome and the same covariates were used as for the previously described linear regression analysis on spherical equivalent. Odds ratios were obtained from the beta regression coefficient using the equation:

$$\ln(\mathrm{OR}) = \frac{\beta}{\mu(1-\mu)}$$

where μ is the fraction of the cases in the sample ($\mu = 0.606$). Genotypes with a minor allele frequency of <0.01 and a minor allele count of <400 were removed from analyses in this group.

23andMe. Participants were all volunteers from the 23andMe personal genomics company customer base. All participants provided informed consent and answered surveys online according to the approved 23andMe human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com). The participants were identified as myopia cases if they self-reported a diagnosis of myopia or reported suffering from symptoms of myopia (see Supplementary Notes for more detail).

DNA extraction and genotyping were performed on saliva samples by Clinical Laboraroty Improvement Amendments (CLIA)-certified and College of American Pathologists (CAP)-accredited clinical laboratories of Laboratory Corporation of America. Samples were genotyped on one of four genotyping platforms and batches (Illumina HumanHap550, BeadChip, SNPs and Illumina OmniExpress, as well as a variable number of custom SNP assays). Only samples with more than 98.5% genotyping success rate were included. Ethnic categorization was conducted using an SVM model which classified individual haplotypes into one of the 31 reference populations derived from public data sets (the Human Genome Diversity Project, HapMap and 1000 Genomes), as well as 23andMe customers who have reported having four grandparents from the same country. Genotypes were imputed against the September 2013 release of 1000 Genomes Phase1 reference haplotypes using a Beagle haplotype graph-based phasing algorithm for the autosomal loci and Minimac2 (ref. [67]) for X Chromosome loci.

Association test results were computed by linear regression assuming additive allelic effects using imputed allele dosages. Covariates for age, gender and the first ten principal components to account for residual population structure were also included into the model.

The Genetic Epidemiology Research in Adult Health and Aging cohort. GERA is part of the Kaiser Permanente Research Program on Genes, Environment, and Health and has been described in detail elsewhere[68]. It comprises adult male and female consenting members of Kaiser Permanente Northern California (KPNC), an integrated health care delivery system, with ongoing longitudinal records from vision examinations. For this analysis 34,998 adults (25 years and older), who self-reported as non-Hispanic white, and who had at least one assessment of spherical equivalent obtained between 2008 and 2014, were included. All study procedures were approved by the Institutional Review Board of the Kaiser Foundation Research Institute. Participants underwent vision examinations, and most subjects had multiple measures for both eyes. Spherical equivalent was assessed as the sphere + cylinder/2. The spherical equivalent was selected from the first documented assessment, and the mean of both eyes was used. Individuals with histories of cataract surgery (in either eye), refractive surgery, keratitis or corneal diseases were excluded from further analyses.

DNA samples from GERA individuals were extracted from Oragene kits (DNA Genotek) at KPNC and were genotyped at the Genomics Core Facility of the University of California, San Francisco. DNA samples were genotyped using the Affymetrix Axiom arrays (Affymetrix). SNPs with an initial genotyping call rate of ≥97%, an allele frequency difference of ≤0.15 between males and females for autosomal markers and a genotype concordance rate of >0.75 across duplicate samples were included. In addition, SNPs with genotype call rates of <90% were removed, as were SNPs with a minor allele frequency of <1%.

Imputation pre-phasing of genotypes was done using Shape-IT v2.r72719 (ref. [65]) and variants were imputed from the cosmopolitan 1000 Genomes Project reference panel (phase I integrated release; http://1000genomes.org) using IMPUTE2 v2.3.0 (ref. [69]). Variants with an imputation IMPUTE $r^2 < 0.3$ were excluded, and analyses were restricted to SNPs that had a minor allele count of ≥20.

For each SNP locus, linear regressions of the spherical equivalent for each individual were performed with the following covariates: age at first documented spherical equivalent assessment, sex and genetic principal components using PLINK v1.9 (https://www.cog-genomics.org/plink/1.9). Data from each SNP were modeled using additive dosages to account for the uncertainty of imputation. The top ten ancestry principal components were included as covariates, and the percentage of Ashkenazi ancestry was included to adjust for genetic ancestry, as described previously[68].

The Consortium for Refractive Error And Myopia (CREAM). All participants selected for this study were of European descent, 25 years of age or older. Refractive error, represented by measurements of refraction and spherical equivalent (spherical equivalent = spherical refractive error + cylinder refractive error/2) was the outcome variable for CREAM. Participants with conditions that could alter refraction, such as cataract surgery, laser refractive procedures, retinal detachment surgery, keratoconus, or ocular or systemic syndromes were excluded from the analyses. Recruitment and ascertainment strategies varied by study and were published previously elsewhere[7].

The genotyping process has been described elsewhere[7]. Samples were genotyped on different platforms, and study-specific quality control measures of the genotyped variants were implemented before association analysis was carried out. Genotypes were imputed with the appropriate ancestry-matched reference panel for all cohorts from the 1000 Genomes Project (Phase I v3, March 2012 release). Quality control criteria were used for SNP and sample inclusions. These metrics were similar to those described in a previous GWAS analysis and detailed information for each cohort is described elsewhere[7].

To avert sample overlap, cohorts from the United Kingdom (1985BBC, ALSPAC-Mothers, EPIC-Norfolk, ORCADES and Twins UK) were excluded from the GWAS meta-analysis. Association analyses were performed as described elsewhere[7]. For each individual cohort, a single-marker analysis for the phenotype of SphE (in diopters) was carried out with linear regression with adjustment for age, sex and up to the first five principal components. For all non-family-based cohorts, one of each pair of relatives was removed. In family-based cohorts, mixed-model-based tests of association were used to adjust for within-family relatedness.

The European Prospective Investigation into Cancer (EPIC) study. The EPIC-Norfolk study is one of the UK arms of a broad pan-European prospective cohort study designed to investigate the etiology of major chronic diseases[70,71]. This

study was conducted following the principles of the Declaration of Helsinki and the Research Governance Framework for Health and Social Care. The study was approved by the Norfolk Local Research Ethics Committee (05/Q0101/191) and East Norfolk & Waveney NHS Research Governance Committee (2005EC07L). All participants gave written, informed consent. Refractive error was measured in both eyes using a Humphrey Auto-Refractor 500 (Humphrey Instruments). Spherical equivalent was calculated as spherical refractive error plus half the cylindrical error for each eye.

The EPIC-Norfolk participants were genotyped using the Affymetrix UK Biobank Axiom Array (the same array as was used in the UK Biobank study); 7,117 participants contributed to the current study. SNP exclusion criteria included: call rate < 95%, abnormal cluster pattern on visual inspection, plate batch effect evident by significant variation in minor allele frequency and/or Hardy–Weinberg equilibrium $P < 10^{-7}$. Sample exclusion criteria included: DishQC < 0.82 (poor fluorescence signal contrast), sex discordance, sample call rate < 97%, heterozygosity outliers (calculated separately for SNPs with minor allele frequency of >1% or <1%), rare allele count outlier and impossible identity-by-descent values. Individuals with relatedness corresponding to third-degree relatives or closer across all genotyped participants were also removed from further analyses. After these steps were taken all participants were of European descent. Data were pre-phased using SHAPEIT[65] v2 and were imputed to the Phase 3 build of the 1000 Genomes project[72] (October 2014) using IMPUTE[69] v2.3.2.

The relationship between allele dosage and mean spherical equivalent was analyzed using linear regression adjusted for age, sex and the first five principal components. Analyses were carried out using SNPTEST v2.5.1.

Statistical analyses. We conducted two meta-analyses. For the initial meta-analysis (discovery), we used summary statistic results from the UK Biobank 1st and 2nd subset, the GERA and 23andMe studies.

For the final meta-analysis, we used all available information (UK Biobank 1st and 2nd subsets, the GERA, 23andMe and CREAM Consortium studies).

For all meta-analyses we applied a *z*-score method, weighted by the effective population sample size, as implemented in METAL[73]. No genomic control adjustment was applied during the meta-analysis.

The effective population size was calculated per individual locus and was equal to the total sample size if a linear regression or linear mixed model were used. For the case–control study the effective population was calculated as:

$$N.\text{eff} = 2/\left(\frac{1}{N.\text{cases}} + \frac{1}{N.\text{controls}}\right)$$

as recommended before[74], where *N*.eff is the effective sample size, *N*.cases is the number of cases considered to have myopia and *N*.controls is the number of subjects considered not to have myopia. Following this method, we calculated that for the full-sample analysis of 542,934 subjects, due to the presence of two case–control cohorts, our effective sample size was 379,227.

Only SNPs with minor allele frequency of at least 1%, which were available from at least 70% of the maximum number of participants across all studies and that were not missing in more than one strata (cohort), were considered further.

Conditional analyses were conducted using the conditional and joint analysis on summary data (COJO) as implemented in the GCTA program[75] to identify independent effects within associated loci as well as to calculate the phenotypic variance explained[76] by all polymorphisms associated with the trait after the conditional analyses. The threshold of significance was set at 5×10^{-8} and the collinearity threshold was set at $r^2 = 0.9$.

Genomic inflation was assessed using the package 'gap' in R (https://cran.r-project.org). To distinguish between the effect of polygenicity and those effects arising from sample stratification or uncontrolled population admixture, the LD score regression intercepts were calculated using the program LD Score (https://github.com/bulik/ldsc).

Bivariate genetic correlations between refractive error and other complex traits whose summary statistics are publicly available were assessed following previously described methodologies[77], using the program LD Score.

To assess the potential value of the loci that are associated with refractive error to predict myopia, regression-based models were trained and tested separately in two groups. The training set comprises the European UK Biobank participants for whom the spherical equivalent measurements were available. The models were tested in the EPIC-Norfolk cohort, which was not part of any of the analyses through which the genetic associations were identified.

The model included age, sex and the major genetic variants associated with refractive error after the conditional analysis. Three different definitions of myopia were used based on sliding spherical equivalent thresholds: $M_1 \leq -0.75$ D, $M_2 \leq -3.00$ D and $M_3 \leq -5.00$ D. These three different definitions of myopia were chosen to correspond to the generally accepted definitions of 'any myopia', 'moderate myopia' and 'high myopia'. For the latter, we opted for the −5.00 D threshold, because definitions based on the more stringent threshold of ≤−6.00 D would have not allowed for a sufficient number of cases in the testing set. For the purpose of these analyses, a 'control' was any subject who did not have myopia, defined by a mean spherical equivalent of ≥−0.5 D.

A receiver operating characteristic curve was drawn for each case and the AUC was calculated. The R programming language and software environment for statistical computing was used for both the logistic regression models ('glm') and to evaluate the performance of the model ('ROCR').

Polymorphisms associated at a GWAS level ($P < 5 \times 10^{-8}$) were clustered within an 'associated genomic region', defined as a contiguous genomic region where GWAS-significant markers were within 1 Mb of each other. Significant polymorphisms were annotated with the gene inside whose transcript-coding region they were located, or alternatively, if located between two genes, with the gene nearest to them. The associated genomic regions were collectively annotated with the gene that was overlapping, or nearest to the most significantly associated variant within that region.

The known relationships between identified genetic loci and other phenotypic traits were derived from two data sets: OMIM, which is a continuously curated catalog of human genes and the phenotypic changes that result from their polymorphic forms and the GWAS Catalog[45] which is a curated catalog of previous GWAS association of SNPs or genes with other phenotypic traits.

The R package MendelianRandomization v3.4.4 was used for Mendelian randomization analyses.

Disease-relevant tissues and cell types were identified by analyzing gene expression data together with summary statistics from the meta-analysis of refractive error in all five cohorts, as described elsewhere[41]. Expression data were obtained from the following sources: 1) the GTEx release v7 (https://gtexportal.org/home/datasets), 2) Fetal and adult corneal, trabecular meshwork and ciliary body RNA sequencing data described previously[39] and 3) data from the subset of subjects with presumed healthy adult retinas (age-related macular degenration, AMD classification grade 1) from data sets described elsewhere[40].

As the transcription data were heterogeneous and were expressed in various units, expression levels for all tissues were rank-transformed. Hierarchical clustering by means of the 'hclust' package in R was used to help visualize similarities and differences in patterns of transcript expression across different tissues.

SMR uses GWAS variants as instrumental variables and gene expression levels or methylation levels as mediating traits, to test whether the causal effect of a specific variant on the phenotype-of-interest acts via a specific gene[42]. The SMR tests were performed using three different reference databases; the summary statistics of eQTL associations in the untransformed peripheral blood samples of 5,311 subjects[43], as well as statistics from the eQTL effects and *cis*-methylation analyses (*cis*-mQTL), both of which were carried out in brain tissues[44].

The gene-set enrichment analysis was implemented in the MAGENTA software[78]. We used the versions from September 2017.

Results of three statistical tests for natural selection were imported from the 1000 Genomes Selection Browser[79].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Summary statistics from the cohorts that participated in the meta-analysis can be downloaded from ftp://twinr-ftp.kcl.ac.uk/Refractive_Error_MetaAnalysis_2020 and public repositories such as the GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics). These freely downloadable summary statistics are calculated using all cohorts described in this manuscript, except for the 23andMe participants. This is due to a non-negotiable clause in the 23andMe data transfer agreement, intended to protect the privacy of the 23andMe research participants. To fully recreate our meta-analytic results, all bona fide researchers can obtain the 23andMe summary statistics by emailing 23andMe (dataset-request@23andme.com) and subsequently meta-analyzing them along with the freely accessible summary statistics for all the other cohorts.

References

64. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
65. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
66. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
67. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
68. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).
69. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
70. Riboli, E. & Kaaks, R. The EPIC Project: rationale and study design. European prospective investigation into cancer and nutrition. *Int. J. Epidemiol.* **26** (Suppl 1), S6–S14 (1997).

71. Hayat, S. A. et al. Cohort profile: a prospective cohort study of objective physical and cognitive capability and visual health in an ageing population of men and women in Norfolk (EPIC-Norfolk 3). *Int. J. Epidemiol.* **43**, 1063–1072 (2014).
72. Delaneau, O., Marchini, J. & The 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
73. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
74. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
75. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
76. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
77. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
78. Segre, A. V. et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
79. Pybus, M. et al. 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res.* **42**, D903–D909 (2014).

Acknowledgements

P.T.K. and P.J.F oversaw the UK Biobank eye data acquisition with support from the National Institute for Health Research (NIHR), Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. The UK Biobank Eye and Vision Consortium was supported by grants from UK NIHR (BRC3_026), Moorfields Eye Charity (ST 15 11 E), Fight for Sight (1507/1508), The Macular Society, The International Glaucoma Association (IGA, Ashford UK) and Alcon Research Institute. V.V. is supported by a core UK Medical Research Council (MRC) grant MC_UU_00007/10. 23andMe thanks research participants and employees of 23andMe for making this work possible (a list of contributing staff is available in the Supplementary Note). Genotyping of the GERA cohort was funded by the US National Institute on Aging, the National Institute of Mental Health and the National Institute of Health Common Fund (RC2 AG036607); data analyses were funded by the National Eye Institute (NEI R01 EY027004, E.J.) and the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK116738, E.J.). The CREAM GWAS meta-analysis was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant 648268 to C.C.W.K), the Netherlands Organisation for Scientific Research (NWO, 91815655 to C.C.W.K) and the National Eye Institute (R01EY020483). V.J.M.V. acknowledges funding from the Netherlands Organisation for Scientific Research (NWO, grant 91617076). S.M. acknowledges support from the National Health and Medical Research Council (NHMRC) of Australia (grants 1150144, 1116360, 1154543, 1121979). EPIC-Norfolk infrastructure and core functions are supported by the MRC (G1000143) and Cancer Research UK (C864/A14136). Genotyping was funded by the MRC (MC_PC_13048). A.K.P. is supported by a Moorfields Eye Charity grant. P.J.F. received support from the Richard Desmond Charitable Trust, the National Institute for Health Research to Moorfields Eye Hospital and the Biomedical Research Centre for Ophthalmology. RW and PGH were supported by the National Eye Institute of the National Institutes of Health under award number R21EY029309. M.J.S. is a recipient of a Fight for Sight PhD studentship. K.P. is a recipient of a Fight for Sight PhD studentship. P.G.H. is the recipient of a FfS ECI fellowship. P.G.H. and C.J.H. acknowledge the TFC Frost Charitable Trust Support for the KCL Department of Ophthalmology. Statistical analyses were run in King's College London on the Rosalind HPC LINUX Clusters and cloud servers. The UK Biobank data were accessed as part of the UK Biobank projects 669 and 17615. J.S.R. is supported in part by the NIHR Biomedical Research Centres at Moorfields Eye Hospital and the UCL Institute of Ophthalmology, and at the UCL Institute of Child Health and Great Ormond Street Hospital, and is an NIHR Senior Investigator. P.M.C. was funded by the Ulverscroft Foundation. O.A.M is supported by Wellcome Trust grant 206619_Z_17_Z and the NIHR Biomedical Research Centre at Moorfields Eye Hospital and the UCL Institute of Ophthalmology.

Author contributions

P.G.H., J.S.R., E.J. and C.J.H. conceived and designed the study. P.T.K., P.J.F. and J.S.R. contributed to the collection of data. P.G.H., H.C., A.P.K., R.W., M.S.T., J.Y., K.K.T., P.M.C., V.V., J.A.G and E.J. performed statistical analysis. A.P.K., M.J.S., K.P., K.K.T., AS and J.A.G. performed post-GWAS follow-up analyses. P.G.H., H.C., A.P.K., R.W., J.S.R., E.J. and C.J.H. wrote the manuscript with help from O.A.M., P.M.C., R.B.M., V.J.M.V., A.S., R.A.S., N.W., A.W.H., D.A.M., C.C.W.K., S.M., P.T.K., P.J.F. and J.A.G. who helped with the interpretation of the results.

Competing interests

23andMe is a consumer genomics company.

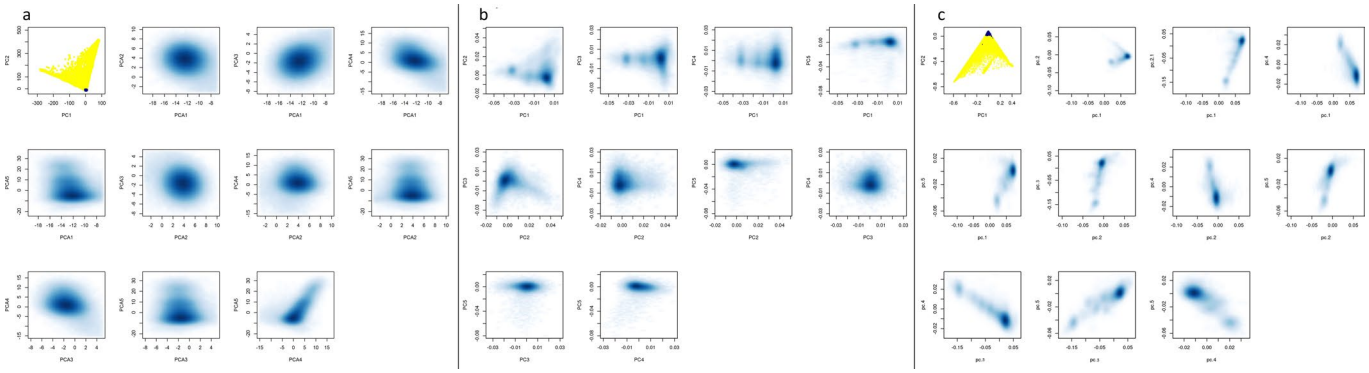
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0599-0.

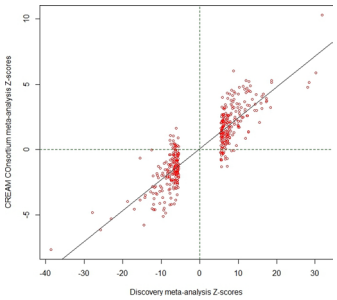
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0599-0.

Correspondence and requests for materials should be addressed to P.G.H.

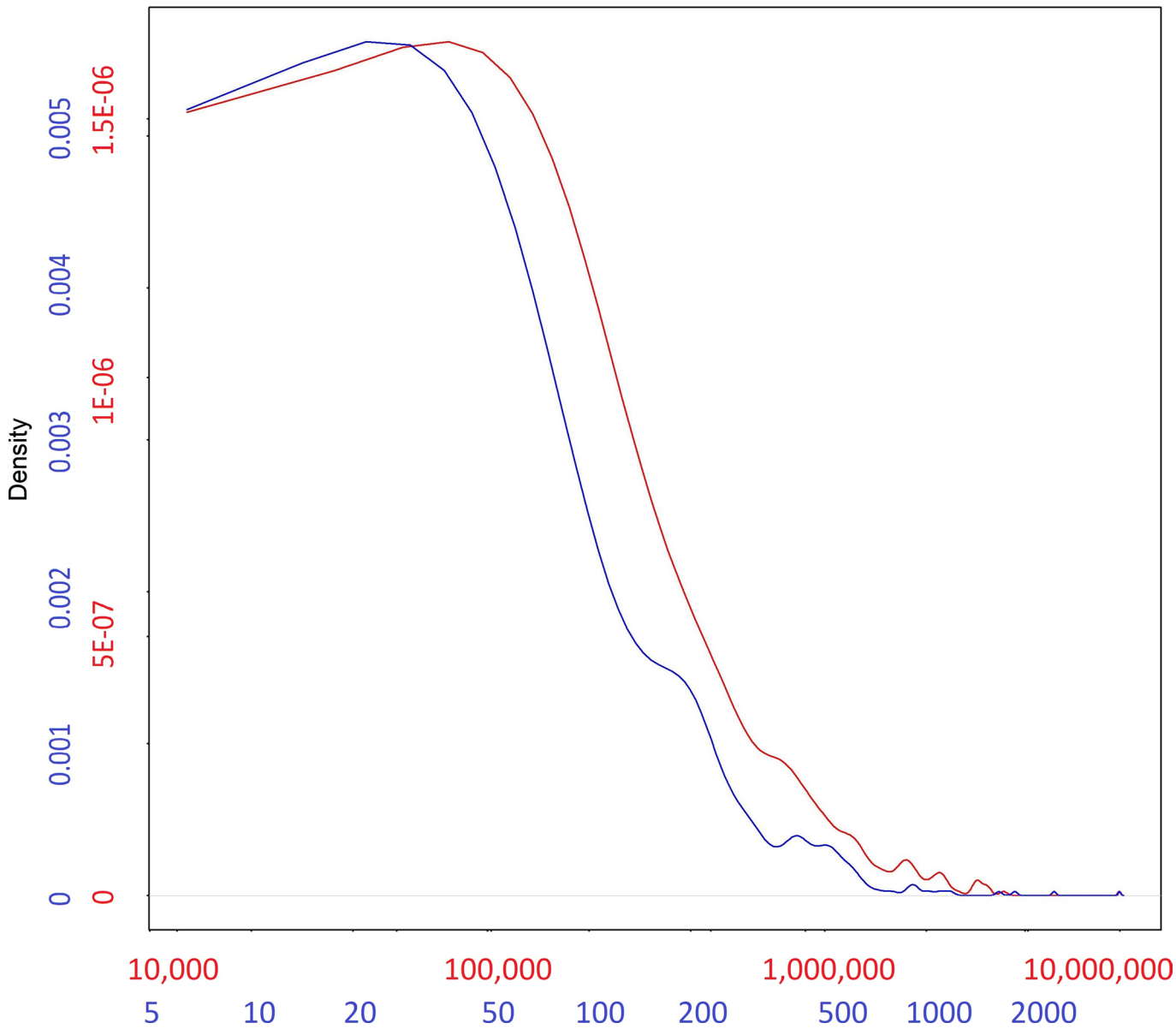
Reprints and permissions information is available at www.nature.com/reprints.



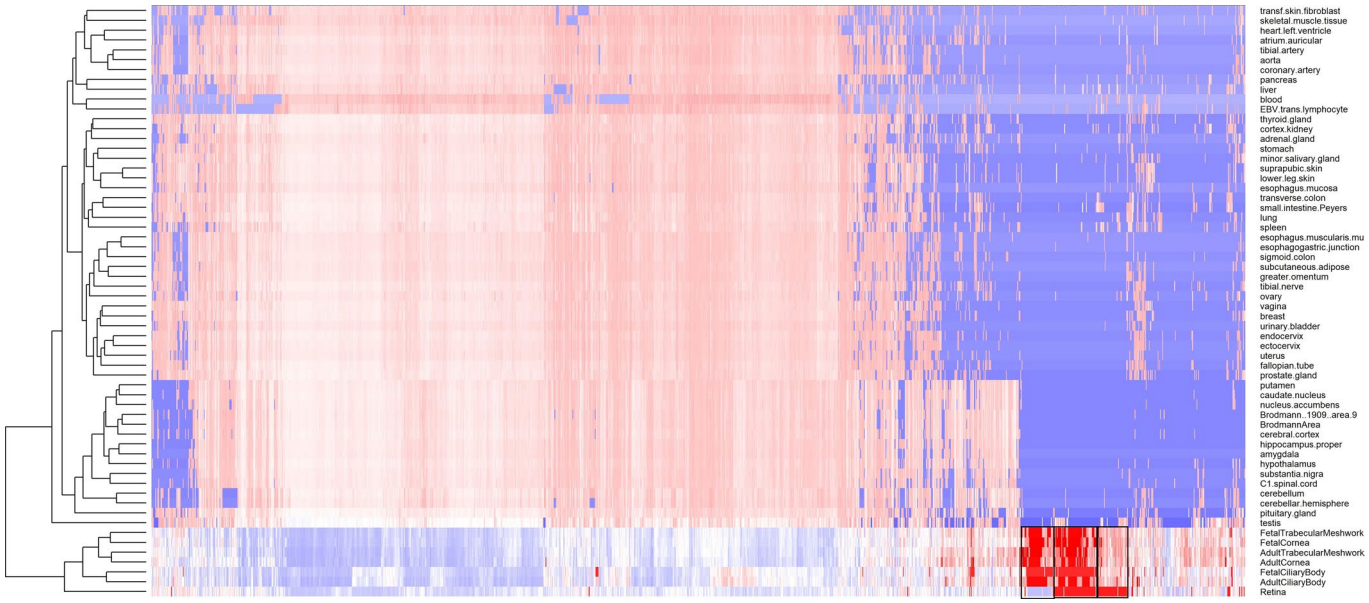
Extended Data Fig. 1 | Principal components plots of the subjects in the main participating cohorts. a) UK Biobank (including the 102,117 subjects with direct refraction measurement and the imputed 108,956 likely myopes to 70,941 likely non-myopes, for a total of 179,897 subjects), **b**) Genetic Epidemiology Research on Adult Health and Aging (GERA, N = 34,998), **c**) 23andMe (106,086 cases and 85,757 controls, or 191,843 subjects in total).



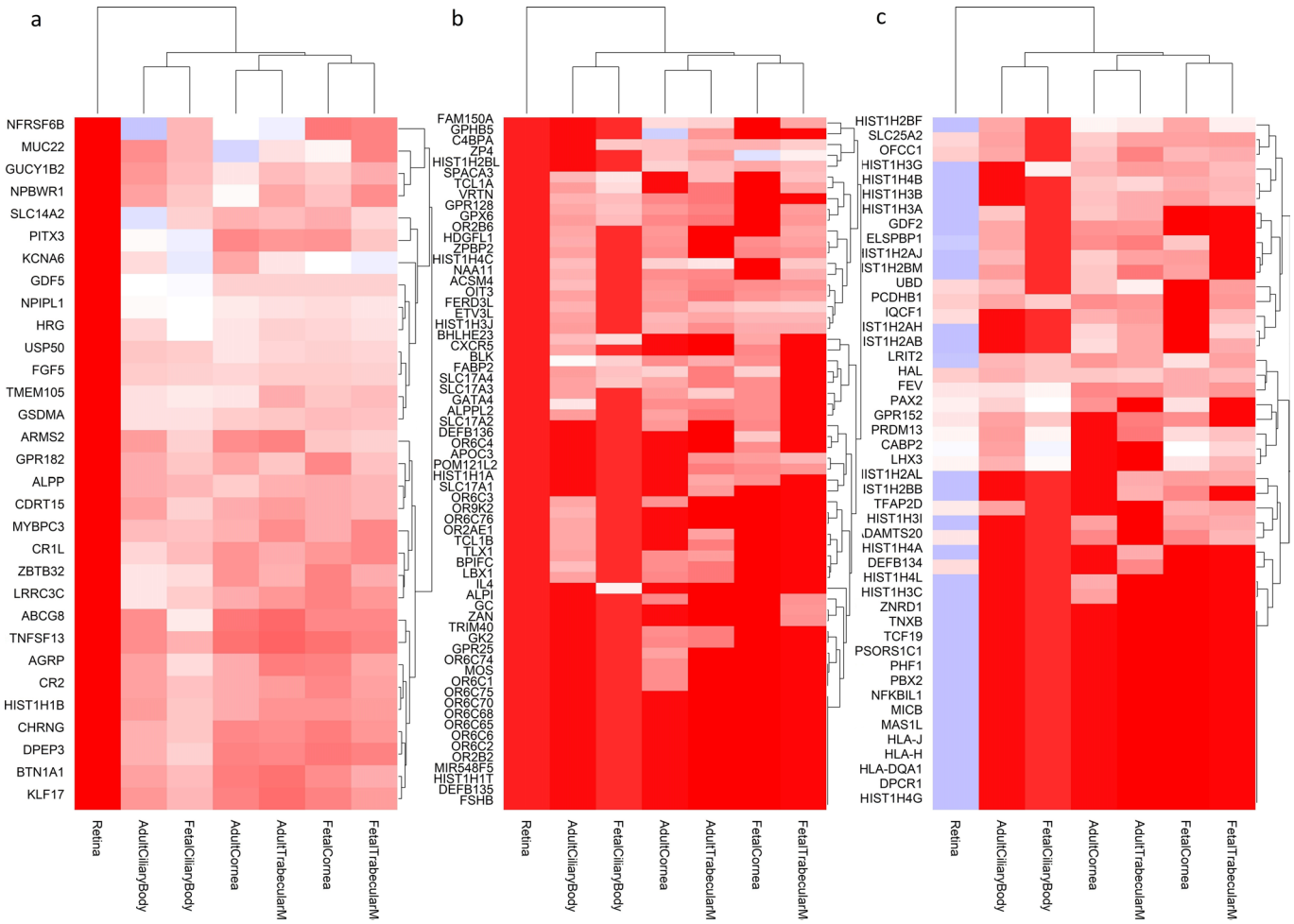
Extended Data Fig. 2 | Correlation of effect sizes between the discovery cohort meta-analysis. Effect sizes are from two analyses, discovery (UK Biobank analysis on spherical equivalent + GERA, spherical equivalent + 23andMe, self-reported myopia cases and controls + UK Biobank inferred myopia cases and controls, for a total of N = 508,855 subjects) and the replication from the non-British CREAM Consortium participants (N = 34,079), used as replication. The z-scores for the discovery are on the y-axis and those from the CREAM cohort in the x-axis.



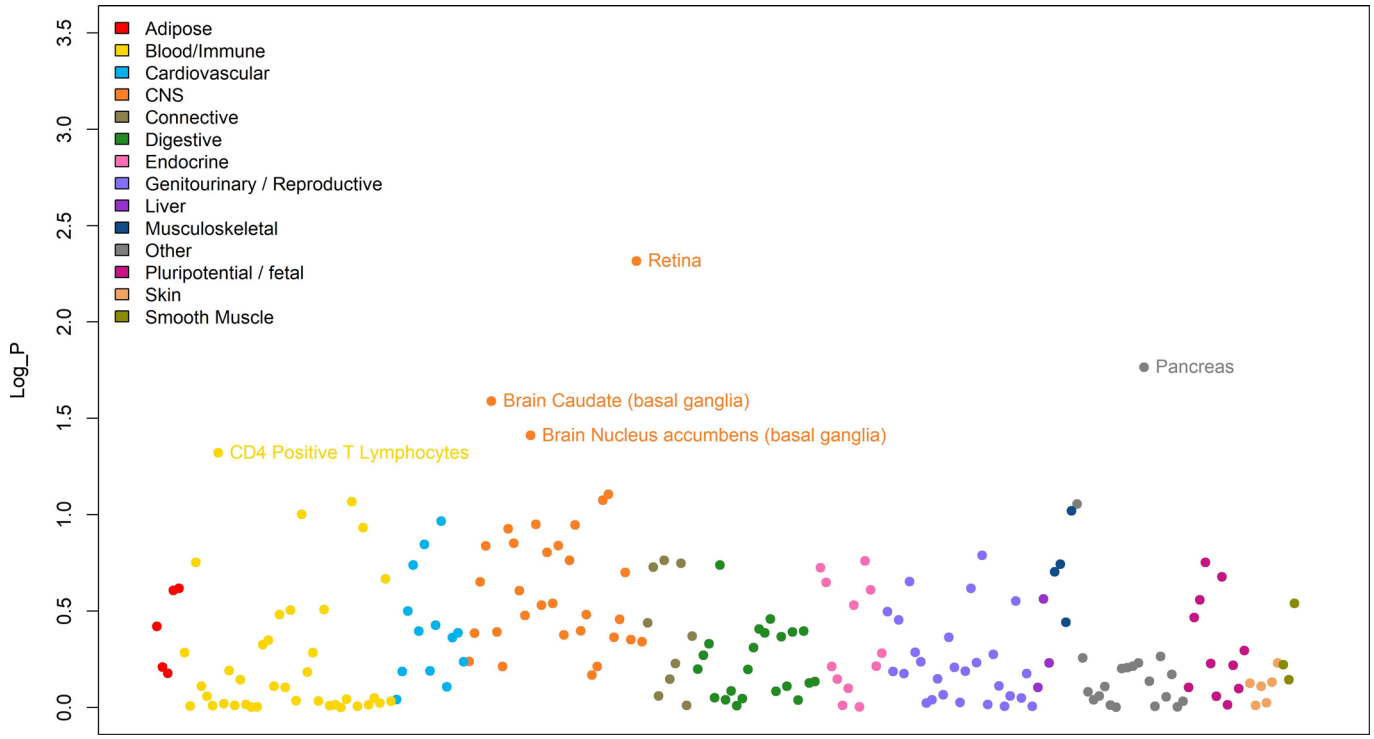
Extended Data Fig. 3 | Distribution of the base-pair length (red) of the 449 regions associated in the meta-analysis of all available cohorts (from Supplementary Table 3), alongside the distribution of number of SNPs (blue) for each region. Numbers in each of the axes in the figure are differentially colored to match the density curve they correspond to: red for the length of the region and blue for the number of SNPs.



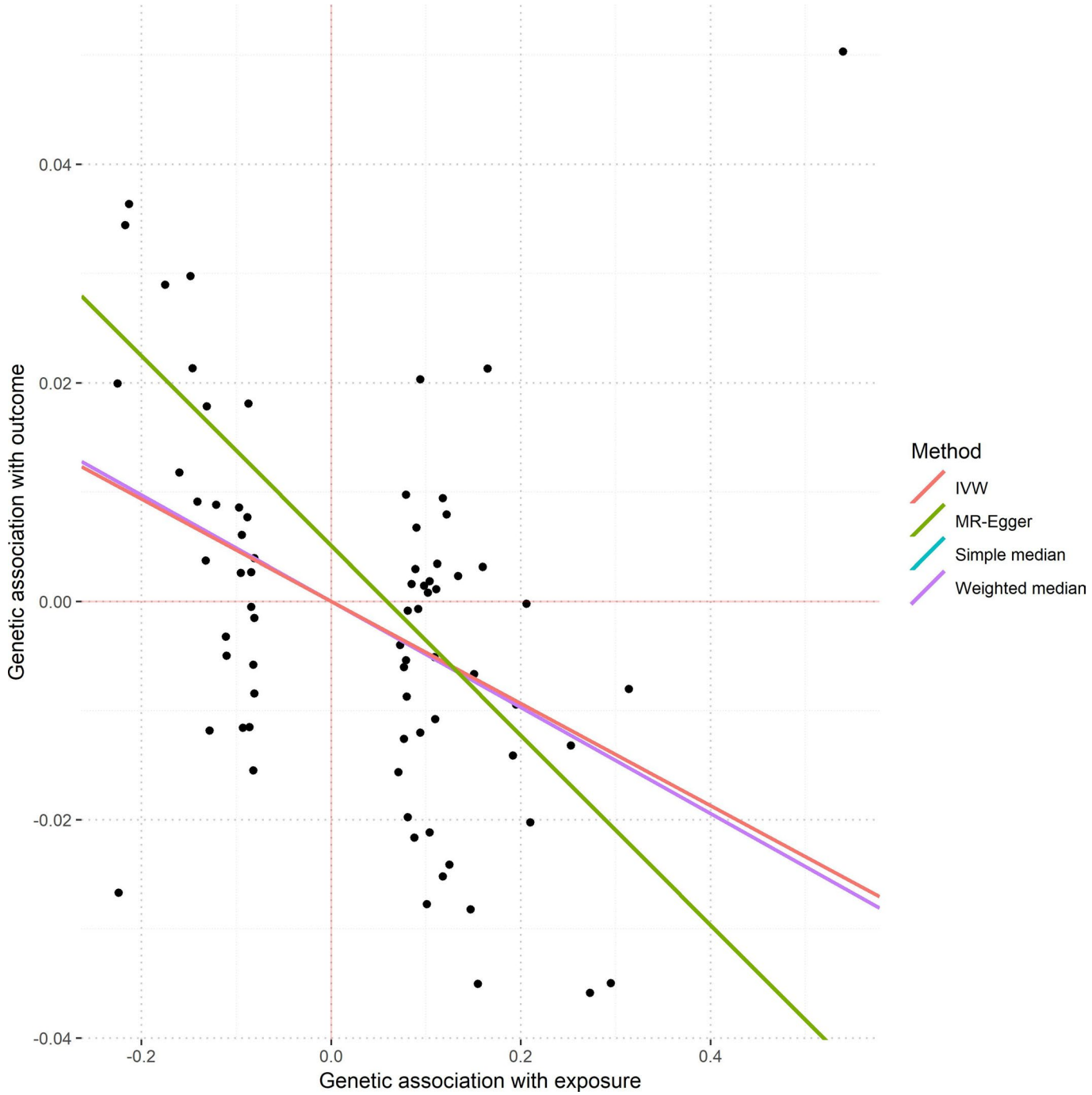
Extended Data Fig. 4 | Expression of genes located in the associated loci (from Supplementary Table 3) along the x-axis, across several human body tissues (y-axis). The colors represent the centile ranking of the expression level of the gene in the tissue of interest. The hotter colors represent higher ranking of the gene expression and the colder colors low expression. Both genes and tissues are clustered in accordance with their pattern similarity. The symbol of all the genes could not be visualized and therefore are removed for the sake of clarity. Eye tissues, whether fetal or adult, appear to have similar patterns of gene expressions (clustered together at the bottom of the figure). Genes that are highly expressed in eye tissues fall in three clusters, shown with a black box. These clusters are displayed in more detail in Figure 4A, B and C.



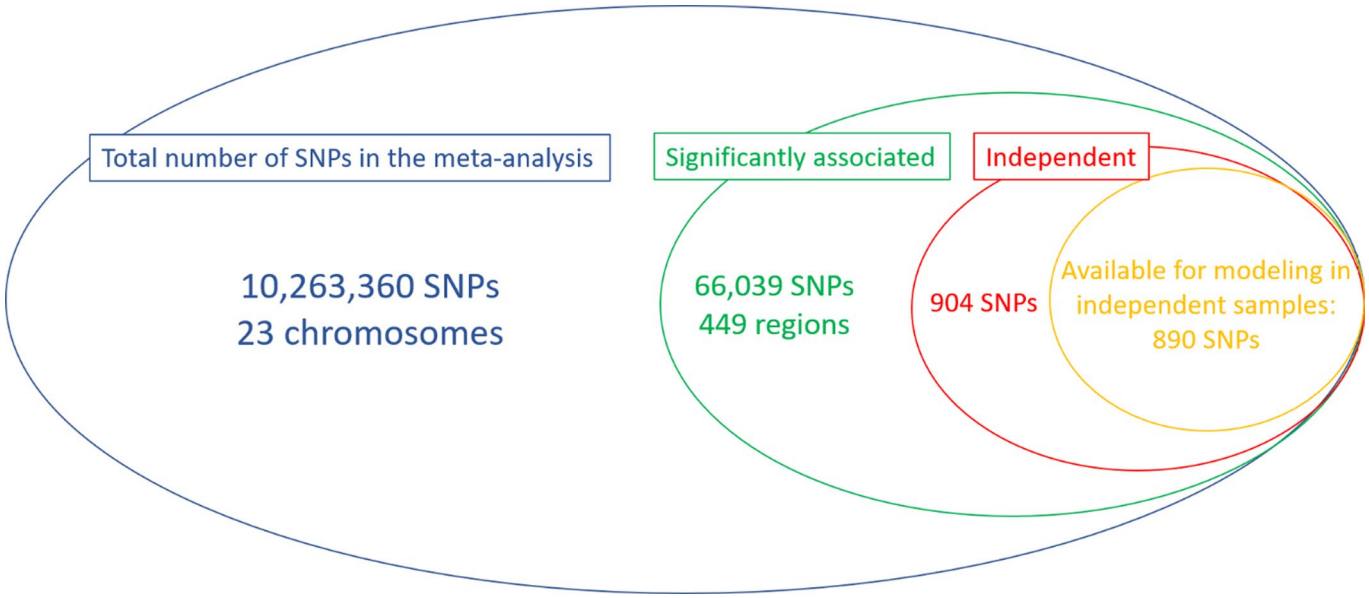
Extended Data Fig. 5 | Genes from the regions associated with RE (from Supplementary Table 3) that are particularly expressed in eye tissues, compared to non-ocular tissues. These clusters are those highlighted in Supplementary Figure 3, but for the sake of clarity they are shown in transposed orientation compared to the previous figure (here genes in the y-axis and eye tissues in the x-axis), but same color codes as before. The dendrograms represent the degree of similarity observed for both tissues and gene expressions. The clusters are given in the order in which they were clustered together, from left to right: **a**) genes that are expressed more in other ocular tissues (fetal and adult) but much less in the adult retina. **b**) genes that are highly expressed in the retina and other ocular tissues, and **c**) genes that are expressed in the retina, but less in the other ocular tissues tested.



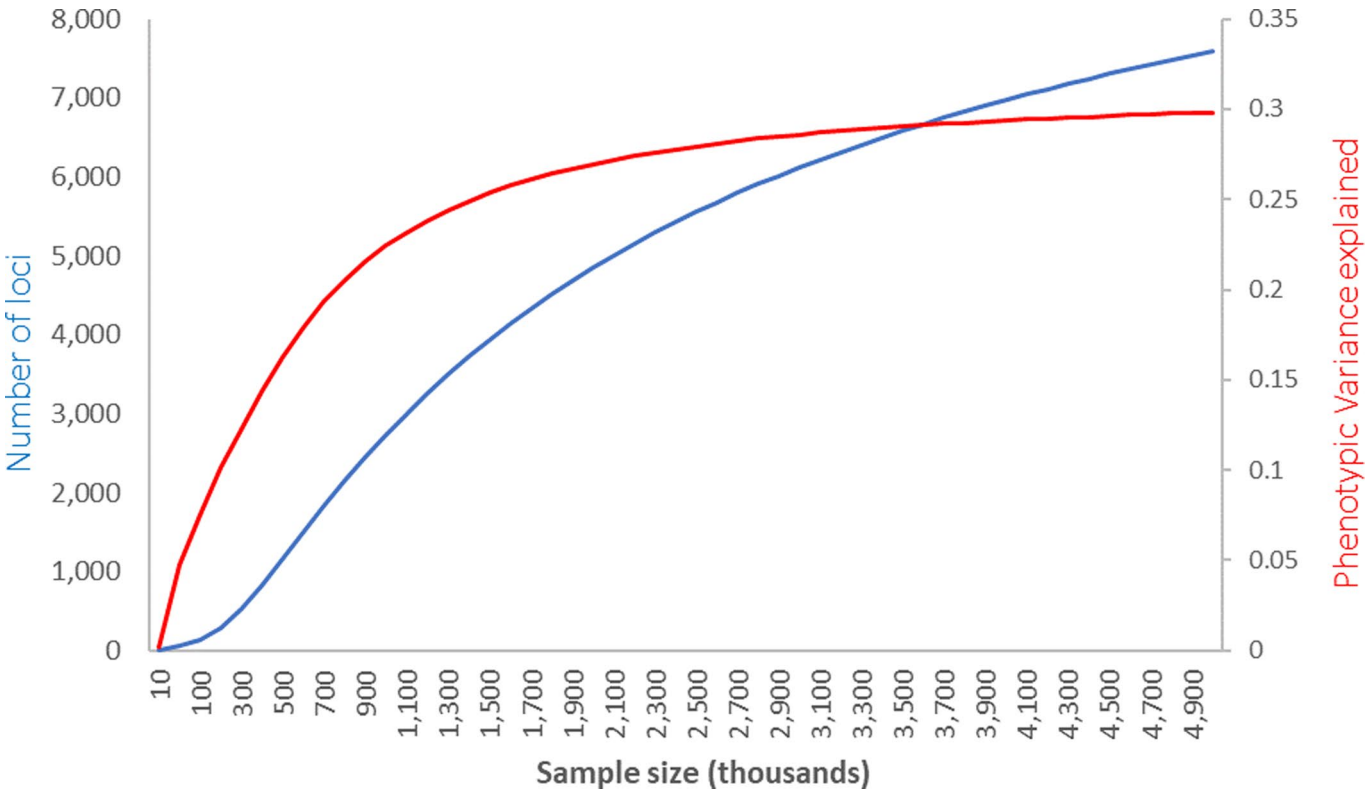
Extended Data Fig. 6 | Results of the LD score regression analysis applied to specifically expressed genes on multiple tissue for the meta-analysis results. Each point represents one tissue or cell line (along the x-axis) and the log10 value of the p-value for the enrichment of the meta-analysis results among genes expressed in these tissues. There were 205 tests carried out, one in each tissue and cell line, therefore only tissues with a correlation p-value< 0.00025 (Log_P > 3.6 in this figure), would have been significant after multiple testing. This condition was not fulfilled for any of the available tissues.



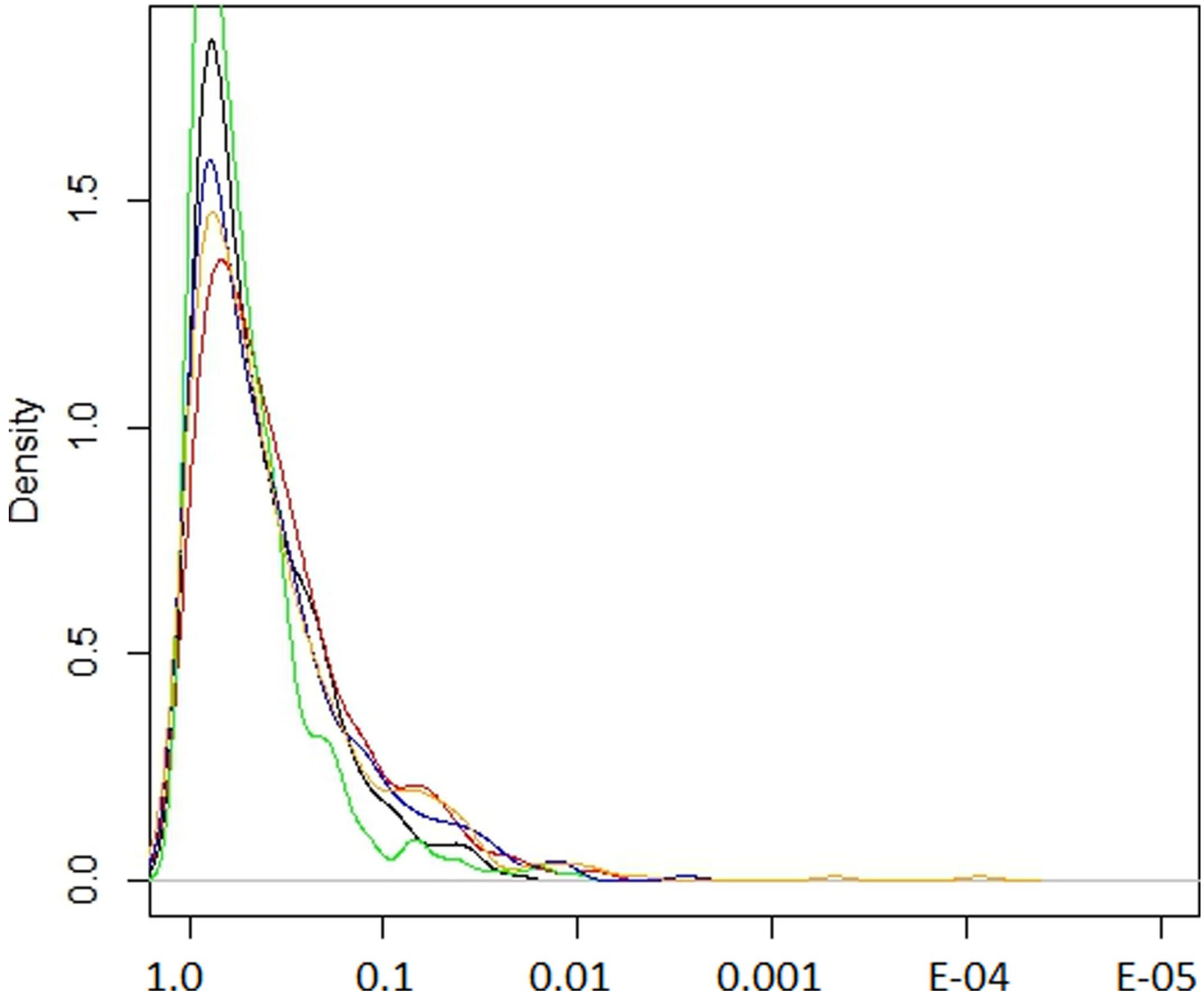
Extended Data Fig. 7 | Mendelian randomization results on causality of IOP over refractive error. Single points in the graph represent coordinates determined by the effect of each specific SNP over IOP (x-axis, mmHg) and spherical equivalent (y-axis, Diopter units). A total of 73 SNPs associated with IOP, but not directly associated with refractive error (that is $p > 0.05$) were selected as instruments. Values of associations with IOP were obtained from a meta-analysis of 139,555 European participants (Reference [50] in the manuscript) and the refractive error associations from 102,117 UK Biobank subjects. The lines represent the regression lines from each model, as specified in the figure legend. In some cases, these lines may not visible because they overlap (please refer to the values underneath the figure).



Extended Data Fig. 8 | Venn's Diagram of the number of SNPs considered in each of the stages of this study. The different circles represent various stages, inclusion in the meta-analysis (blue), identification of significant loci (green), conditional analysis results identifying independent effects (red) and the total number of SNPs available for inclusion in prediction and heritability estimation in the independent (that is not part of the original meta-analysis) EPIC-Norfolk cohort (orange).



Extended Data Fig. 9 | Prediction for the total number of SNPs and phenotypic variance explained as a function of GWAS sample size in future studies, based on the distribution of effects observed in the current meta-analysis. The plot lines show the predicted relationship between the number of loci associated with refractive error (left vertical axis, blue line) and the variance they help explain (red line, right vertical axis), as a function of the sample size (x-axis) used in future GWAS or meta-analyses. These projections are consistent with the observed results, where an effective sample of 379,227 identified 904 independent signals after a conditional analysis, explaining 12–16% of refractive error variability.



Extended Data Fig. 10 | The distribution of various natural selection test scores for SNPs associated with refractive error. The values on the x-axis represent the ranking in terms of natural selection observed and the y-axis the density of that rank. The different tests shown are iHS, XP-EHH (CEU vs YRI), XP-EHH average score, XP-EHH maximum score and Tajima scores (black, green, red, blue and yellow respectively).

nature research

Corresponding author(s): Pirro G. Hysi

Last updated by author(s): Jan 15, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For most cohorts, data were obtained through volunteer-based recruitment, weather population based (UK Biobank, the CREAM consortium participants), patient population (GERA) or customer based questionanires (23andMe). For a subset of the UK Biobank participants, an approximate phenotypic status was obtained through a support vector machine (SVM) analyses of correlated variagles (primarily age when the participants' vision was first corrected and demographics). The e1071 package for R was used.
Data analysis	All analyses were conducted using open source software, as specified in the Online Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

UK Biobank individual level data is publicly available to bona fide researchers (https://www.ukbiobank.ac.uk/using-the-resource/).
Data from the GERA cohort is accessible through the NIH dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v2.p2)and additional information may be provided upon request.
Association summary statistics data from the 23andMe may be provided upon request (https://research.23andme.com/collaborate/).
Summary statistics data can be obtained as per the original publication (Tedja et al. Nature Genetics 50.6 (2018): 834-848.)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study was conducted on all refractive error information globally available for individuals of European descent. No target power, effect, or sample size applies to thsi study.
Data exclusions	To minimize confounding effects arising from ethnicity-related heterogeneity, we excluded all samples that were not a full European ancestry, as routinely recommended in similar genetic association studies. In addition, we removed data from participants whose other medical conditions or episodes could have altered refractive error-related measurements or refractive error state (for example eye surgery, eye infections, etc.)
Replication	Reproducibility of results was assured by comparing association results, effect size and direction as well as significance of association tests among several population-based cohorts.
Randomization	No randomization procedure was needed.
Blinding	No blinding was needed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Regression-based association analyses had refractive error as the outcome, number of alleles at each polymorphic locus as predictors, making adjustment for age, sex and the most important principal components.
Recruitment	Different participating studies followed different recruitment strategies: the UK Biobank has recruited 500,000 volunteers among the adult population of the United Kingdom; the GERA study includes consenting patients from the Kaiser Permanente North California; the 23andMe cohort includes consenting participants from the 23andMe customer base; the CEAM consortium participants were recruited among the general population in several countries from many academic institutions, as described elsewhere (Tedja et al. Nature Genetics 50.6 (2018): 834-848.)
Ethics oversight	This study was conducted in accordance with the principles of the Declaration of Helsinki. Written and informed consent was obtained from participants in each of the studies prior to the inclusion of their genetic, demographic or clinical data in any of the analyses. Approval was obtained for each of the cohorts by ethical committees: North-West Research Ethics Committee (ref 06/MRE08/65) for the UK Biobank, Institutional Review Board of the Kaiser Foundation Research Institute for the GERA study, Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com) for 23andMe, Norfolk Local Research Ethics Committee (05/Q0101/191) and East Norfolk & Waveney NHS Research Governance Committee (2005EC07L) for the EPIC-Norfolk study, and competent national or regional ethics governance authorities with jurisdiction over the areas of recruitment of the individual studies participating in the CREAM cohort.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Not applicable (not a clinical study)
Study protocol	N/A
Data collection	N/A
Outcomes	N/A